DOCUMENT RESUME

ED 276 338 HE 019 834

AUTHOR Bender, Timothy A.
TITLE Classroom Feedback and Students' Ability Level.
PUB DATE Apr 86
NOTE 16p.; Paper presented at the Annual Meeting of the American Educational Research Association (67th, San Francisco, April 16-20, 1986).
PUB TYPE Reports - Research/Technical (143) -- Speeches/Conference Papers (150)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Ability Grouping; *Answer Keys; Educational Research; Error Patterns; *Feedback; Higher Education; *Multiple Choice Tests; *Review (Reexamination); Undergraduate Students

ABSTRACT

The effect that student ability level has on receiving feedback following classroom tests was studied. Forty-four undergraduates enrolled in four educational psychology classes were assigned to low or high ability groups based on their total score from the first four exams. Two classes were trained in a feedback technique, and the remaining two classes served as the control. One class meeting following each exam was used as a feedback session. All students were given their scored answer sheets and a copy of the exam and asked to review their exam. Students were told to first review items they answered incorrectly and to search the text and their notes for the correct answer. They were then to review items they answered correctly and to review the text concerning items about which they were uncertain. Students in the control sessions were told to review their exams until they were satisfied. All students were administered the same multiple-choice semester tests and the final, which consisted of 40 repeated items, 20 verbatim and 20 paraphrased items, and 10 new items. Only data concerning the 40 repeated items were analyzed. Attention was directed to: the number of correct responses; types of errors for the verbatim and paraphrased items; and new, perseverative, and different error patterns. Findings are discussed. (SW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED276338

Classroom Feedback and Students' Ability Level

Timothy A. Bender

Southwest Missouri State University

Presented in a round table session at the annual meeting of the American Educational Research Association in San Francisco, April 1986.

HE 019 834

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☐ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Timothy A. Bender

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Classroom Feedback and Students' Ability Level

Feedback following classroom tests should afford students the opportunity to learn from their mistakes. Kulhavy, White, Topp, Chan, & Adams (1985) suggested that feedback corrects inaccurate information. However, Bender (1984) suggests that feedback should serve the three functions of confirming correct responses, disconfirming incorrect responses, and finally correcting inaccurate information. Bender indicated that feedback would only serve these functions if students processed the feedback effectively. Both Kulhavy et al. (1985) and Bender (1984) were based on the assumption that feedback acts as a source of information, the effectiveness of which is dependent on the processing given to the information.

Given a pretest-posttest examination system, three error patterns can occur when students fail to effectively process feedback (Phye, Gugliemela, & Sola, 1976). A 'new error' occurs when feedback is not used to confirm a previously correct response. A 'different error' occurs when feedback disconfirms, but does not correct an initial error. Finally, a 'perseverative error' occurs when feedback is not used to disconfirm an initial error. Figure 1 illustrates a model of the

---

Insert Figure One About Here

---

processing of feedback. New errors would occur when the processing breaks down on the left half of the model. Different errors would occur when students learn their initial answer was incorrect, then the processing breaks down. This would occur on the right in Figure 1. Perseverative errors occur when students do not process that their initial answer was incorrect.

Hunt (1978) indicated that higher ability students may use more effective information processing. If the effect of feedback is dependent on the effectiveness of the information processing, it should be possible to demonstrate differences in the processing of feedback between higher and lower ability students. Bender (1984) discussed an examination review technique which appeared to improve the use of feedback in lower ability students. However, there were a number of problems with that research. The subjects were few in

number and from a very small private liberal arts college for women. The posttest was a verbatim posttest. Finally, subjects served as their own controls. This study is an attempt to replicate Bender's earlier findings while using a greater number of students from a larger university. Students were given verbatim and paraphrased posttest items and separate classes were used for a control.

Method

Subjects

Subjects were 56 undergraduates enrolled in four educational psychology classes all taught by the same professor. Subjects were assigned to the low or high ability groups on the basis of their total score from the first four exams. This assignment was determined after the semester and grading were completed. Due to attendance problems and in an effort to keep the conditions balanced, only 44 subjects were used in the final analysis, 11 in each condition.

Procedure

Two classes were trained in the feedback technique and the remaining two classes served as the control. One class meeting following each exam was used as a feedback session. All students were given their scored answer sheets and a copy of the exam and asked to review their exam. The answer key was displayed by the use of an overhead projector.

In the classes which received the feedback training, students were told to first review those items they answered incorrectly and search the text and their notes for the correct answer. Next they were to review those items they answered correctly and review the text concerning those items for which they were uncertain. Students returned their exams and answer sheets once they were finished with the review. Students in the control sessions were only told to review their exams until they were satisfied, then to return them.

All subjects were administered the same four 50-point multiple-choice semester tests, and the same 50-point final. Five items from each of the semester tests were repeated verbatim on the final. Five more items from each of the semester tests were paraphrased

on the final. The remaining ten items were over material not previously tested. Thus, the final was comprised of 40 repeated items, 20 verbatim and 20 paraphrased and 10 new items. Only data concerning the 40 repeated items were analyzed.

Subjects were determined as being either of high or low ability on the basis of the total points from the first four exams. Thus, four conditions were formed; high ability treatment, low ability treatment, high ability control, and low ability control.

Differences between ability groups and treatment groups were expected. It was hypothesized that higher ability control subjects would answer more items correctly and commit a lower proportion of new, different, and perseverative errors than the lower ability control subjects. It was also expected that the higher ability control subjects would correct a greater proportion initial errors than would the lower ability control subjects. In keeping with the results of Bender (1984), no differences for error pattern were expected between the performances of the different ability treatment groups.

Results

Three separate analyses were completed. The first was an analysis of the number of correct responses. The other analyses comprised an error analysis of the types of errors committed for the verbatim and paraphrased items. The first error analysis was for for the proportion of corrected items. The second was for new, perseverative, and different error patterns.

Number Correct

A 2 (low versus high ability) x 2 (treatment versus control) x 4(item group: combined pretest verbatim score, combined pretest paraphrased score, final verbatim score, final paraphrased score) mixed factor ANOVA, with the last factor treated as a within-subjects variable was used to analyze general performance. Tukey's Honestly Significant Differences (HSD) test was used to make comparisons between means in any interactions. The dependent measure, number correct in each item group, included the number correct from the items which were to be repeated verbatim from the regular semester tests, number correct from the items which were to be paraphrased from the regular

semester tests, number of verbatim final items correct, and number of paraphrased final items correct.

The between-subjects main effect of ability was significant, $F(1,40) = 109.883$, $p < .00001$, with mean scores across treatment and sessions for the low and high ability subjects of 9.955 and 15.227, respectively. The maximum score would be 20.

The two-way interaction of ability and treatment was significant $F(1,40) = 7.347$, $p < .01$. Means for this interaction can be found in Table One. Low

---

Insert Table One About Here

---

ability subjects who receive feedback instructions performed more poorly on the repeated items than did either high ability group, HSD = 4.73, n = 11, $p < .01$. Low ability control subjects performed more poorly than did high ability treatment subjects, HSD = 4.73, n = 11, $p < .01$, and high ability control subjects, HSD = 3.81, n = 11, $p < .05$.

The two-way interaction between ability and item group was also significant, $F(3,120) = 5.498$, $p < .002$. Means for this interaction can be found in Table Two.

Insert Table Two About Here

High ability subjects performed better on all item groups than did low ability subjects, HSD = 2.49, n = 22, $p < .01$. No differences were found within ability levels.

Error Analysis

The dependent measures for the error analyses included the proportions of new, perseverative, different, and corrected errors for the verbatim and paraphrased items. The proportion of new errors was determined by dividing the number of items which were answered correctly on the first test but incorrectly on the second by the total number correct on the first. The proportion of perseverative errors was determined by dividing the number of items answered incorrectly in the same manner on both the pretests and postest by the number of items answered incorrectly on the pretest.

The proportion of different errors was determined by dividing the number of items answered incorrectly on both the pretests and the posttest, but with different incorrect answers, by the number incorrect on the pretest. The proportion of corrected errors was determined by dividing the number of items answered incorrectly on the pretests, but corrected on the posttest, by the number answered incorrectly on the pretest. All proportions were transformed using an arcsin transformation before analysis (Kirk, 1968), and are reported as transformed scores.

A 2 (low versus high ability) x 2 (instructions versus control) x 2 (corrected errors on verbatim items versus corrected errors on paraphrased items) mixed factor ANOVA, with the last factor treated as a within subjects variable was completed. No significant main effects or interactions were found.

A 2 (low versus high ability) x 2 (instructions versus control) x 6 (error pattern: new verbatim, perseverative verbatim, different verbatim, new paraphrased, perseverative paraphrased, different paraphrased) mixed factor ANOVA, with the last factor treated as a within subjects variable was also completed. Comparisons between means were completed using Tukey's HSD test.

A significant between-subjects main effect for ability was found, $F(1,40) = 22.606$, $p < .0001$, with mean transformed proportions of errors for low and high ability subjects of 1.136 and .803, respectively.

A significant within-subjects main effect for error patterns was found, $F(5,200) = 2.855$, $p < .02$. The transformed proportions for these patterns can be found in Table Three. Across treatment and ability

---

Insert Table Three About Here

---

levels, subjects committed a greater proportion of new errors with the paraphrased items than different errors with the verbatim items, HSD = .321, n = 44, $p < .05$.

The two-way interaction between error pattern and ability level was significant, $F(5.200) = 2.27$, $p < .05$. Means for this interaction can be found in Table Four. Low ability subjects committed a greater

Insert Table Four About Here

proportion of new errors with verbatim items than different errors, HSD = .597, n = 22, $p < .05$. Low ability subjects also committed a greater proportion of new errors with verbatim items than the proportions of errors committed by high ability subjects in the categories of new or different errors with verbatim items, perseverative or different errors with paraphrased items, HSD = .597, n = 22, $p < .01$, and perseverative errors with verbatim items, HSD = .521, n = 22, $p < .05$. Low ability subjects also committed a greater proportion of new errors with the paraphrased items than the high ability subjects' proportion of different errors with the paraphrased items, HSD = .597, n = 22, $p < .01$, and the high ability subjects' proportion of new errors with the verbatim items, HSD = .521, n = 22, $p < .05$. Finally, low ability subjects committed a greater proportion of perseverative errors with the verbatim items than the high ability subjects' proportions of new errors with the verbatim items and different errors with the paraphrased items, HSD = .521, n = 22, $p < .05$.

Discussion

The expected differences for ability were partially supported. However, instead of the differences in number correct and error patterns being limited to the control groups, they appeared across treatment conditions. Low ability subjects answered fewer items correctly in all the item groups. Low ability subjects also performed poorly with respect to new errors with both the verbatim and paraphrased items and perseverative errors with the verbatim items. The most frequent error pattern across ability groups was new errors with the paraphrased items, while the least frequent error pattern was different items with the verbatim items.

The only difference within an ability level and item type appeared for low ability subjects and verbatim items. Low ability subjects committed a greater proportion of new errors than different errors. No differences between the proportions of error types were found for the high ability subjects.

The only difference within error pattern and type of item appeared for new errors on verbatim items. Low ability subjects committed a greater proportion of new errors than did high ability subjects. No differences in the proportions of error types appeared between ability levels for paraphrased items. It appears that the low ability students were not effective in processing the feedback concerning their initially correct answers.

The remaining differences in the error pattern and ability interaction indicate that high ability subjects committed a very low proportion of new errors with verbatim items and a low proportion of different errors with paraphrased items. Low ability subjects tended to commit a relatively large proportion of new errors in both item types and a large proportion of perseverative errors in verbatim items. .

These results support the assumption that feedback serves as a source of information, the effectiveness of which depends on how the information is used. This information should be used to confirm previously correct responses as well as to correct inaccurately encoded information. The finding of a greater proportion of error types for the low ability subjects indicates that they are not as proficient at using feedback as the higher ability subjects. The ability differences also provide information about how feedback may be used by students in the classroom.

New errors are expected when subjects fail to use the feedback to confirm initially correct items. This is a reinforcing function. It appears that this reinforcing function does not occur as well for the lower ability subjects as for the high ability subjects. Apparently the reinforcer, i.e., feedback is not commanding the low ability learners' attention; or lower ability subjects have not developed effective strategies for processing the information in classroom feedback.

Perseverative and different errors should be examined together. According to the feedback model of Bender (1984), perseverative errors occur when subjects do not make any use of feedback to learn from their initial errors. Different errors occur when subjects learn only which alternative is incorrect, but not which is correct. Therefore, a subject who commits a high proportion of perseverative errors may produce a low proportion of different errors. If feedback is being somewhat effectively processed, no differences

between the proportions of these error types would be expected. Within each ability level, it appears that subjects are somewhat proficient at using feedback concerning initial errors. However, low ability subjects did have some difficulty with perseverative errors on the verbatim items. This also suggests lower ability students may not have the same strategies for using classroom feedback as do the higher ability students.

All students appeared to use the feedback to affect their representations of the course information to some extent. However, higher ability students appeared to be more adept at this. If feedback were used simply to memorize correct responses, you would expect to find a greater proportion of perseverative and new errors on the paraphrased items than on the verbatim items. In both of these cases, the student who simply memorizes responses without comprehending the content of the response would not be able to identify the memorized response on a paraphrased retention test. Higher ability students profited more than the lower ability students from the confirmatory function of feedback. Higher ability students also used the feedback to learn when they were incorrect, but did not use the feedback situation to full advantage. This is evident in the lack of differences between ability groups in the proportions of different and corrected errors.

Teachers may profit from this line of research if it can be demonstrated that the differences in how feedback is used is consistent for identifiable groups of students, such as higher and lower ability students. Once consistent differences are found, the next step is to develop techniques which promote the more effective use of classroom feedback in students. Apparently, a procedure which simply provides guidelines for the use of feedback and then asks the students to follow the guidelines is not consistently effective.

References

Bender, T. (1984). Improving the use of classroom feedback. Presented at the annual meeting of the American Educational Research Association, New Orleans.

Hunt, E. (1978). Mechanics of verbal ability. Psychological Review, 85, 109-130.

Kirk, R. E. (1968). Experimental Design: Procedures for the behavioral sciences. Belmont, CA: Brooks/Cole Publishing Company.

Kulhavy, R., White, M., Topp, B., Chan, A., & Adams, J. (1985). Feedback complexity and corrective efficiency. Contemporary Educational Psychology, 10, 285-291.

Phye, G., Gugliemella, J., & Sola, J. (1976). Effects of delayed retention on multiple-choice test performance. Contemporary Educational Psychology, 1, 26-36.

Table 1. Mean Number Correct for Ability Level and Treatment.

Ability	Treatment	
	Control	Instructions
Low	10.795_a	9.114_b
High	14.705	15.750

Note. Mean with subscript a was significantly lower than high/control at $p < .05$ and high/instructions at $p < .01$. Mean with subscript b was significantly lower than either high ability condition at $p < .01$.

Table 2. Mean Number Correct for Ability and Type of Item

Item Type	Ability	
	Low	High
Verbatim		
Semester tests	9.909	15.727
Final	9.682	15.773
Paraphrased		
Semester tests	9.227	15.273
Final	11.000	14.136

Note. All high ability means are significantly greater than all low ability means at $p < .01$.

Table 3. Mean Proportions of Error Patterns for Verbatim and Paraphrased Items.

Error Pattern	Item Types	
	Verbatim	Paraphrased
New	1.064	1.144_a
Perseverative	1.055	0.874
Different	0.814	0.866

Note. Means are acrsin transformations. Mean with subscript a is significantly different from mean for different/verbatim at $p < .05$.

Table 4. Mean Proportions of Error Patterns for Item Type and Ability Level

Item Type and Error Pattern	Ability Low	Ability High
Verbatim		
New	1.429_{a}	0.699_{bfh}
Perseverative	1.260_{g}	0.849_{c}
Different	0.831_{b}	0.796_{b}
Paraphrased		
New	1.286_{d}	1.002
Perseverative	0.951	0.797_{b}
Different	1.057	0.675_{beh}

Note. Means are acrsin transformations. Mean with subscript a is significantly greater than means with subscript b at $p < .01$ and mean with subscript c at $p < .05$. Mean with subscript d is significantly greater than mean with subscript e at $p < .01$ and mean with subscript f at $p < .05$. Mean with subscript g is significantly greater than means with subscript h at $p < .05$.

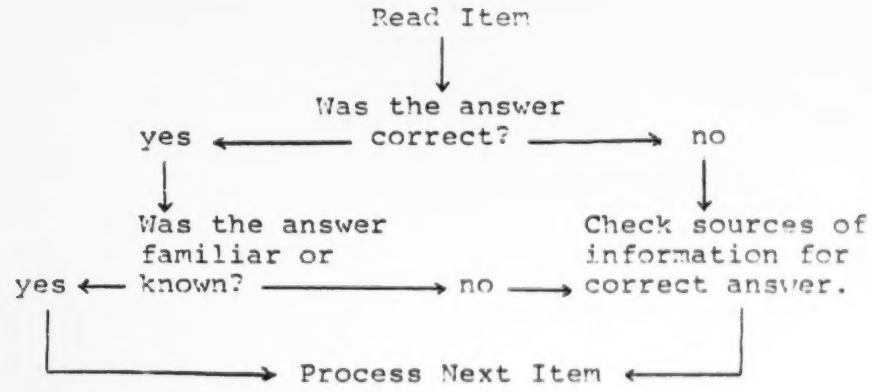

Figure 1. Model for processing exam item feedback.